User-Centric Access to e-Government Information: e-Citizen Discovery of e-Services

Giovanni Maria Sacco

Dipartimento di Informatica, Universita' di Torino
Corso Svizzera 185
10149, Torino, Italy
sacco@di.unito.it

Abstract

Effective and timely user access to public information is one of the most fundamental requirements for e-government and e-democracy. In this context, the discovery of e-services available to citizens plays a very important role, because it represents one of the most frequent points of contact between public administrations and citizens. We present a solution, based on dynamic taxonomies, to end-user discovery of e-services. Differently from mainstream research in semantic web, the solution we propose is intended for the direct use of end-users, rather than for programmatic or agent-mediated access.

Introduction

One of the most overlooked areas in the current debate and research on e-government is e-citizen access to public information. Information available to citizens includes normative material (the official publication of laws and regulations) but also information on the e-services the government offers. Currently, a typical e-government portal provides 100-200 e-services, a number growing as more and more government agencies go online.
Services tend to be perceived by citizens as the friendlier face of government and are often more interesting to the general public than laws and regulations are. However, it is already difficult, and becoming more so, to quickly and effectively find the **right** e-services from a pool of several hundred.
In the following, we briefly review traditional solutions and introduce dynamic taxonomies. Dynamic taxonomies have been proposed as a tool for the universal access of e-government information, from "institutional" information (Sacco, 2005a) to non-normative information such as job-brokering services (Sacco, 2005d). In the present paper, we apply dynamic taxonomies to the end-user discovery of e-government services, a topic that completes the notion of a single, universal access paradigm for all information access needs of e-governments.

Current Solutions

There is no extensive research body on the discovery of e-government services. Fang and Liu Sheng (Fang, 2005) propose a data-mining approach that rearranges service links on the basis of the analysis of web-logs. While this speeds up the most common interactions, it offers no clue to the specific user on the right service for her, and it might well hinder discovery by directing the user to the most common service format. As an example, consider that most governments offer services for minorities: these services would be quite hard to find.
The selection of e-services has traditionally been performed through traditional access paradigms, such as queries on structured database systems, and information retrieval. These methods retrieve data on the basis of precise specifications. However, most search tasks, especially in the present context, are exploratory and imprecise in essence: the user does not usually know precisely what he wants (e.g. a specific service), but rather he needs to explore the information base, find relationships among concepts and thin alternatives out in a guided way.
It is intuitively appealing to consider solutions based on semantic networks (Schmeltz Pedersen, 1993) in the context of ontologies and Semantic Web. This approach is quite expensive in terms of design and maintenance of complex conceptual schemata, and, although more powerful and expressive than plain taxonomies, is more difficult to understand and manipulate by the end user. It is better suited to programmatic access, and, consequently, user interaction must be mediated by specialized agents, which tend to lack transparency and suffer from many of the problems of knowledge-based systems (Brézillon, 1999). The solution we propose in the following can also be used as a complement to complex ontologies: a taxonomy-based model that provides a user-understandable view on complex semantics (Sacco, 2005e).

Dynamic Taxonomies

Dynamic taxonomies (Sacco, 1987, 1998, 2000) are a general knowledge management model for complex, heterogeneous information bases. It has been applied to very diverse areas, including multimedia databases (Sacco, 2004), electronic commerce (Sacco, 2005c), art and museum portals (Yee et al., 2003 and Hyvönen et al., 2004), and medical guidelines (Wollersheim and Rahayu, 2002; Sacco, 2005b). The intension of a dynamic taxonomy is a taxonomy designed by an expert. It does not require any other relationships in addition to subsumptions (e.g., IS-A and PART-OF relationships).

In the extension, items can be freely classified under several topics at any level of abstraction. This multidimensional classification models common real-life situations. First, items are very often about different concepts. Second, items usually have different independent features (e.g. Time, Location, etc.), each of which can be described by an independent taxonomy. These features are often called *perspectives* or *facets*.

In dynamic taxonomies, a concept C is just a label that identifies all the items classified under C. Because of the subsumption relationship between a concept and its descendants, the items classified under C (*items(C)*) are all those items in the *deep extension* of C, i.e. the set of items identified by C includes the *shallow extension* of C (all the items directly classified under C) union the deep extension of C's sons.

There are two important consequences of our approach. First, since concepts identify sets of items, logical operations on concepts can be performed by the corresponding set operations on their extension. Second, dynamic taxonomies can find all the concepts related to a given concept C: these concepts represent the conceptual summary of C. Concept relationships other than subsumptions are inferred through the extension only, according to the following *extensional inference rule*: two concepts A and B are related if there is at least one item d in the infobase which is classified at the same time under A (or under one of A's descendants) and under B (or under one of B's descendants). The extensional inference rule infers relationships on the basis of empirical evidence.

The extensional inference rule can be easily extended to cover the relationship between a given concept C and a concept expressed by an arbitrary subset S of the universe: C is related to S if there is at least one item d in S which is also in items(C). Hence, the extensional inference rule can produce conceptual summaries for any logical combination of concepts and, in addition, for sets of items produced by other retrieval methods such as information retrieval, etc.

Access through Dynamic Taxonomies

The user is initially presented with a tree representation of the initial taxonomy for the entire infobase. Each concept label has also a count of all the items classified under it (i.e. the cardinality of items(C) for all C's). The initial user focus F is the universe (i.e. all the items in the infobase).

In the simplest case, the user can then select a concept C in the taxonomy and *zoom* over it. The zoom operation changes the current state in two ways. First, concept C is used to refine the current focus F, by intersecting it with items(C); items not in the focus are discarded. Second, the tree representation of the taxonomy is modified in order to summarize the new focus. All and only the concepts related to F are retained and the count for each retained concept C' is updated to reflect the number of items in the focus F that are classified under C'. The reduced taxonomy is a conceptual summary of the set of documents identified by F, exactly in the same way as the original taxonomy was a conceptual summary of the universe. In fact, the term *dynamic taxonomy* is used to indicate that the taxonomy can dynamically adapt to the subset of the universe on which the user is focusing, whereas traditional, static taxonomies can only describe the entire universe.

The retrieval process is an iterative thinning of the information base: the user selects a focus, which restricts the information base by discarding all the items not in the current focus. Only the concepts used to classify the items in the focus, and their ancestors, are retained. These concepts, which summarize the current focus, are those and only those concepts that can be used for further refinements. From the human computer interaction point of view, the user is effectively guided to reach his goal, by a clear and consistent listing of all possible alternatives.

Benefits of Dynamic Taxonomies

The advantages of dynamic taxonomies over traditional methods are dramatic in terms of an extremely fast convergence of exploratory patterns and in terms of human factors. Three zoom operations on terminal concepts are sufficient to reduce a 1,000,000-item information base described by a compact taxonomy with 1,000 concepts to an average 10 items (Sacco, 2002). Dynamic taxonomies only require a very light theoretical background: namely, the concept of a taxonomic organization and the zoom operation, which seems to be very quickly understood by end-users. Usability tests on a corpus of art images (Hearst et al. 2002; Yee et al. 2003) showed a significantly better recall than text retrieval and, perhaps more importantly, the feeling that one has actually considered all the alternatives in reaching a result.

Dynamic taxonomies cleanly separate the process of classifying documents from the use of the classification information in the browsing system, and considerably simplify the design of the conceptual taxonomy. First, the extensional inference rule actually performs concept association mining: concept associations, which are often quite dynamic in time, need not be forecasted and accounted for in schema design. In addition, the user is presented with associations the schema designer might not even be aware of.

Second, since dynamic taxonomies synthesize compound concepts, these need usually not be represented explicitly, so that we avoid the exponential growth due to the description of all the possible concept combinations, and the resulting taxonomy is significantly more compact and easier to understand. Sacco (Sacco, 2000) developed a number of guidelines for taxonomies that are compact and easily understood by users. Some are similar to the basic faceted classification scheme by Ranganathan (Ranganathan, 1965): the taxonomy is organized as a set of independent, “orthogonal” subtaxonomies (facets or perspectives). As an example, a compound concept such as *housing for senior citizens* need not be explicitly accounted for, because it can be synthesized from its component concepts: *Service>housing* and *Age>senior*, where Service and Age are facets.

The term *faceted classification systems* sometimes used instead of *dynamic taxonomies* is a misnomer because: a) faceted classification only addresses conceptual modeling and very basic concept composition: conceptual summaries, reduced taxonomies and guided navigation are totally absent, and b) faceted classification is a special case of the multidimensional classification used by dynamic taxonomies.

Discovery of services

Since dynamic taxonomies dynamically reconstruct all the correlations among concepts, breaking compound concepts into their base components allows the user to easily correlate concepts and explore such correlations. In the running example, the user focusing on *Age>senior* will immediately find all the relevant services related to senior citizens (which include *housing*). Most importantly, the interaction is completely symmetrical: the user will find *Age>senior* among the topics related to *Service>Housing* if he starts the other way round. If compound concepts are used (*housing for senior citizens*), correlation information is hidden inside labels, and cannot be carried out automatically but requires the manual inspection of labels. These considerations indicate that the taxonomy for e-services is best organized into a number of facets, tailored according to differences and options in e-services. Two obvious high-level subtaxonomies are services and citizens. Each can be further characterized independently by facets that describe their semantic variety. For instance, if there are specific housing services for senior citizens, seniority must be made visible in the taxonomy, presumably as a descendant of Age. Some obvious facets for citizens are *age*, *sex*, *physical handicaps*, *personal relationships* (e.g. single parents, married couples), etc., while additional facets might be required if specific provisions for some set of citizens exist. An analogous architecture applies to services, with high-level subtaxonomies such as *Housing* (change of address, grants), *Schooling* (enrollments, grants, special schools), etc. will be described.

This simple and easily understood architecture has two important benefits with respect to traditional methods. First, the user is guided to reach her goal no matter what her starting point is. In addition, she is guided to discover all potentially relevant information: for instance, all services offered to senior citizens, or all the subjects that are entitled specific services in housing. No other technique currently offers this level of flexibility.

Second, modification (insertion, deletion, update) of services is especially easy since the interrelations among concepts are dynamically computed.

References

Brézillon, P. 1999. Context in problem solving: A survey, *The Knowledge Engineering Review*, 14(1), pp.1-34.

Fang, X.; Liu Sheng, O. 2005. Designing a better web portal for digital government: a web-mining based approach, Proc. of the dg.o2005 conf. on Digital government research

Hearst, M., et al. 2002. Finding the Flow in Web Site Search, *Comm. of the ACM*, 45, 9

Hyvönen, E.; Saarela, S.; and Viljanen, K. 2004. Application of Ontology Techniques to View-Based Semantic Search and Browsing. Proc. of the First European Semantic Web Symposium (ESWS 2004), *Springer Verlag LNCS 3053*, 92-106

Ranganathan, S. R. 1965. *The Colon Classification*. Rutgers University Press

Sacco, G. M. 1987. Navigating the CD-ROM, Proc. Int. Conf. Business of CD-ROM

Sacco, G. M. 1998. Dynamic taxonomy process for browsing and retrieving information in large heterogeneous data bases, US Patent 6,763,349

Sacco, G. M. 2000. Dynamic Taxonomies: A Model for Large Information Bases. *IEEE Transactions on Knowledge and Data Engineering* 12, 2 , p. 468-479

Sacco, G. M. 2002. Analysis and Validation of Information Access through Mono, Multidimensional and Dynamic Taxonomies, Tech Rep. Univ. of Torino

Sacco, G. M.. 2004. Uniform access to multimedia information bases through dynamic taxonomies, Proc. IEEE Int. Symp. on Multimedia Software Engineering

Sacco, G. M. 2005a. No (e-)Democracy without (e-) Knowledge, TCGOV 2005, *Springer Verlag LNAI* 3416

Sacco, G. M. 2005b. Guided interactive diagnostic systems IEEE Symp. on Computer-Based Medical Systems, Dublin

Sacco, G. M. 2005c. The intelligent e-store: easy interactive product selection and comparison, 7th IEEE Conf. on E-Commerce Technology, IEEE CEC'05

Sacco, G. M. 2005d. Guided interactive information access for e-citizens, EGOV05, *Springer LNCS* 3591

Sacco, G. M. 2005e. Discount semantics: modelling complex data with dynamic taxonomies. Forthcoming.

Schmeltz Pedersen, G. 1993. A Browser for Bibliographic Information Retrieval, Based on an Application of Lattice Theory. Proc. ACM SIGIR Conf.

Wollersheim, D., Rahayu, W. 2002. Methodology For Creating a Sample Subset of Dynamic Taxonomy to Use in Navigating Medical Text Databases. Proc. IDEAS 2002 Conf.

Yee, K-P., et al. 2003. Faceted Metadata for Image Search and Browsing, Proc. ACM CHI Conf.